

## Programme DATA – 2024 : Bourse M2

### Dossier de candidature

par mail à [sandro.bimonte@inrae.fr](mailto:sandro.bimonte@inrae.fr) et [david.sarramia@uca.fr](mailto:david.sarramia@uca.fr)

**TITRE : Conception et implémentation d'un lac de données de robotique agricole**

Pour accompagner la transition agroécologique, les robots ont un rôle essentiel à jouer dans le domaine de l'agriculture intelligente. Ils sont capables d'effectuer des opérations agricoles répétitives et précises sur une longue période avec un faible impact sur l'environnement.

Avec des équipements particuliers, et associés à des technologies d'acquisition et de traitement de données, les robots sont capables d'effectuer de manière autonome des tâches ciblées de manière efficace dans les champs.

De nombreux travaux de recherche portent sur l'agriculture intelligente. Dans le cadre du projet ISITE CAP2025, deux projets se sont intéressés à la gestion de données de l'agriculture. Le premier a réalisé une architecture Big Data pour le traitement des données des robots et des capteurs (avec Apache Kafka et Spark), ainsi qu'une base de données relationnelle pour stocker ces données. Le deuxième de type lac de données, le CEBA (Cloud Environnemental au Bénéfice de l'Auvergne), s'est intéressé à la collecte, l'ingestion et la restitution des données issues de capteurs bas débit en utilisant la pile Elastic et des bases de données relationnelles manipulant des données semi-structurées, sans fonctionnalité d'analyse.

Dans ce travail de stage, nous visons à mettre en place, dans le CEBA, un lac de données (data lake) complet (intégrant les outils des deux projets précédents) pour le stockage et l'exploration des données sources.

En particulier le travail consiste à :

- étudier et comparer les travaux existants sur les lacs de données spatiales : stockage et exploration
- caractériser les flux et implémenter un système d'ingestion (kafka, redpanda...)
- implémenter le système de stockage des données (ex. Apache Hadoop HDFS ou PostGIS)
- concevoir un modèle de métadonnées et implémenter un système de métadonnées
- implémenter un système d'exploration des données (ex. Apache Atlas, Open Metadata, Geonetwork, etc.)

Encadrant : Sandro BIMONTE

Unité de rattachement : TSCF

Date d'obtention de l'HDR : 2015

Etablissement de rattachement : INRAE

Courriel / Téléphone : [sandro.Bimonte@inrae.fr](mailto:sandro.Bimonte@inrae.fr)

Equipe : ROMEA

Co-encadrant : David SARRAMIA

Unités de rattachement : LPC

Etablissements de rattachement : UCA

Courriel / Téléphone [david.sarramia@uca.fr](mailto:david.sarramia@uca.fr)

**Lieu : INRAE, Clermont Ferrand**

**Durée stage 6 mois**

**Début mars avril 2024**

Publications /références associées au sujet :

G. André, B. Bachelet, P. Battistoni, A. Belhassena, S. Bimonte, C. Cariou, F. Chabot, G. Chalhoub, A. Couvent, G. Garani, J. Laneurit, R. Moussa, K. Oikonomou, I. Sammour, M. Sebillio, M. Vilela Souza, N. Tricot, R. Wrembel. LambdaAgroT: a new architecture for agricultural autonomous robots' scheduling: from design to experiments. *Clust. Comput.* 26(5): 2993-3015 (2023)

S. Bimonte, H. Badir, P. Battistoni, H. Bazza, A. Belhassena, C. Cariou, G. Chalhoub, J.C. Corrales, A. Couvent, J. Laneurit, R. Moussa, J.E. Plazas, M. Sebillio, N. Tricot. Data-centric UML profile for agroecology applications: Agricultural autonomous robots monitoring case study. *Comput. Sci. Inf. Syst.* 20(1): 459-489 (2023)

Sarramia, D., Claude, A., Ogereau, F., Mezhoud, J., & Mailhot, G. (2022). CEBA: A Data Lake for Data Sharing and Environmental Monitoring. *Sensors*, 22(7), 2733.

